Achtung!

Dies ist eine Internet-Sonderausgabe des Aufsatzes "Perspektiven des Computereinsatzes in der Orientalistik" von Jost Gippert (1989).
Sie sollte nicht zitiert werden. Zitate sind der Originalausgabe in "Forschungsforum" 2, 1990, 133-136 zu entnehmen.

Attention!

This is a special internet edition of the article "Perspektiven des Computereinsatzes in der Orientalistik" ["Perspectives of the application of computers in Near Eastern studies"] by Jost Gippert (1990).
It should not be quoted as such. For quotations, please refer to the original edition in "Forschungsforum" 2, 1990, 133-136.

Perspektiven des Computereinsatzes in der Orientalistik

Jost Gippert

Nachdem die Domäne des Computers im universitären Einsatz noch bis vor wenigen Jahren im naturwissenschaftlichen Bereich lag, werden elektronische Verfahren heute mehr und mehr auch in den Geisteswissenschaften angewendet. Das primäre Einsatzgebiet liegt dabei zweifellos in der Textverarbeitung; der Computer erweist sich hier als ein universal einsetzbares Hilfsmittel, das die Gestaltung eines Textes von seiner Konzipierung bis zur Drucklegung in der Hand des Autors ermöglicht und herkömmlichen Verfahren somit überlegen ist.

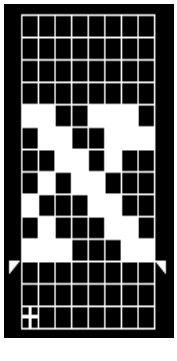
Als entscheidende Voraussetzung für einen weitergehenden Siegeszug des Computers in den Geisteswissenschaften ist die Verfügbarkeit von Zeichen und Schriften anzusehen, die über den Grundvorrat des lateinischen Alphabets hinausgehen. Während im Bereich von Großrechenanlagen und im internationalen Datentransfer noch immer ein Zeichenvorrat von nur 128 Zeichen verwendet wird, der in der sog. ASCII-Norm die 26 Buchstaben des Lateinalphabets in Groß- und Kleinschreibung, die Ziffern von 0 bis 9, die wichtigsten Satzzeichen sowie 32 Steuercodes für Zeilenschaltungen u.ä. umfaßt, war schon die Aufstockung auf 256 Zeichen, mit denen heute jeder Personal Computer nach dem IBM-System operiert, ein bedeutender Fortschritt. Dieser erweiterte Zeichensatz ist auf die west- und nordeuropäischen Nationalalphabete deutsch, französisch, italienisch, spanisch, niederländisch und schwedisch zugeschnitten und umfaßt über die 128 ASCII-Zeichen hinaus z.B. diakritische Kombinationen wie *ä, ö, ü, á, é, å* etc., die Ligatur *æ*, die Währungssymbole für das engl. Pfund (£) oder den holl. Gulden (*ƒ*), einige in der Mathematik gebrauchte Symbole wie z.B. ∩ oder ≡, einige - ebenfalls nach den Bedürfnissen der Mathematik ausgewählte - griechische Buchstaben wie z.B. α oder Σ sowie sog. Grafikzeichen wie z.B. │ oder ╜, die zur Erzeugung von Linien oder Rahmen am Bildschirm gebraucht werden. Als ein Kuriosum ist festzuhalten, daß das deutsche "scharfe" *ß* in diesem Zeichensatz fehlt; stattdessen muß das Zeichen β verwendet werden, das im System das griech. *beta* repräsentiert.

Es liegt auf der Hand, daß für einen universalen Einsatz in den Geisteswissenschaften auch dieser erweiterte Zeichensatz noch keine geeignete Grundlage abgibt. Weder ein Slavist, der z.B. tschechisches Wortmaterial zu verarbeiten hat, noch ein Klassischer Philologe, der das Altgriechische nicht transkribieren will, findet sämtliche von ihm benötigten Zeichen wieder; von den Bedürfnissen eines Orientalisten, der mit Schriften wie Hebräisch oder Arabisch arbeiten muß, ist dabei ganz zu schweigen. Diese Probleme sind innerhalb der letzten fünf Jahre an verschiedenen Orten und mit unterschiedlichen Verfahren angegangen worden, und zufriedenstellende Lösungen liegen bereits in großem Umfang vor. Das betrifft zunächst die Erweiterung des Zeichenvorrats um weitere lateinschriftliche Sonderzeichen und diakritische Kombinationen, dann die Verfügbarkeit von kompletten, auf das Altgriechische zugeschnittenen griechischen Zeichensätzen einschließlich der zahlreichen Akzentkombinationen sowie das kyrillische Alphabet, wie es heute im Russischen und den anderen slavischen Nationalsprachen verwendet wird. Im Bereich der Orientalistik bleibt jedoch noch genügend Entwicklungsarbeit zu leisten, bis die hier anfallenden Sprachen und Schriften dem Benutzer eines Computers mit demselben Komfort zur Verfügung stehen. Was bei einer solchen Entwicklungsarbeit zu bedenken ist, soll im folgenden kurz umrissen werden.

Die Anforderungen, die an ein mit Originalschriften operierendes Textverarbeitungssystem gestellt werden, müssen sich prinzipiell an dem heute erreichten Leistungsstandard der "normalen" lateinschriftlichen Textverarbeitung orientieren. Das bedeutet zunächst, daß die betreffende Schrift am Bildschirm sichtbar sein muß und, möglichst in verschiedenen Schriftgrößen und -ausgestaltungen, auch auf dem Drucker ausgegeben werden kann, ferner, daß sie in einer adäquaten Weise, im Normalfall über die Tastatur, erzeugbar sein muß. Es bedeutet weiter, daß die Standardfunktionen der Textverarbeitung wie z.B. automatisches Suchen und Ersetzen bestimmter Zeichen oder Zeichenfolgen, Fuß- und Endnotenverwaltung, Zeilenfunktionen wie Blocksatz, Zentrieren oder rechtsbündige Ausrichtung, Spalten- und Tabellenerstellung in der Fremdschrift mit derselben Effektivität ausgeführt werden können wie in der Lateinschrift. Wünschenswert ist darüber hinaus eine eindeutige Kodierung, die es ermöglicht, den erzeugten Text von einem System auf ein anderes zu übertragen; dies ist z.B. die Voraussetzung für die Weitergabe an eine Druckerei ("Drucken von Diskette"). Eine unabdingbare Anforderung, die die wissenschaftliche Arbeit stellt, ist letztlich die gleichzeitige Verfügbarkeit von Originalschrift(en) und Lateinschrift innerhalb desselben Texts.

Um diesen Erfordernissen gerecht zu werden, bedarf es in der Regel eines vielschichtigen Systems von ineinandergreifenden Programmelementen, sog. Treibereinheiten, die von der internen Struktur des verwendeten Rechners und seiner Zusatzgeräte abhängen. Das betrifft zum einen die Erzeugung von Schriftzeichen auf Bildschirm und Drucker, die vorrangig auf sogenannten Bitmatrizen, d.h. Rastern aus Einzelpunkten, basiert. Umfang und Dichte dieser Raster hängen einerseits von der Leistungsfähigkeit der verwendeten Geräte ab, andererseits von der gewünschten Größe der Zeichen. Eine beispielhafte Zusammenstellung ist den folgenden Abbildungen zu entnehmen, wo der hebräische Buchstabe Aleph in fünf Matrizes dargestellt ist: Abbildung 1 zeigt das Aleph in einer typischen Matrix für die Bildschirmausgabe,

Abb. 1: Bitmatrix für Bildschirmzeichensatz



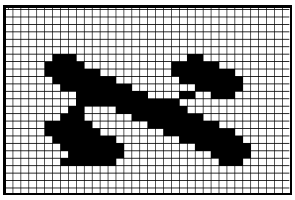
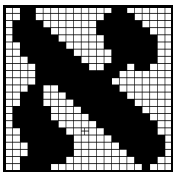
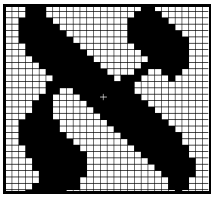


Abb. 2: Bitmatrix für 24-Nadel-Drucker (24 · 36 Punkte)

die nach dem sogenannten "EGA" - Standard ("Enhanced Graphics Adapter") 14 · 8 Punkte umfaßt. Abbildung 2 zeigt das Aleph in der Bitmatrix für 24-Nadel-Drucker, die im Schönschriftmodus üblicherweise mit einer Matrix von 24 · 36 Punkten operieren. Abbildungen 3 bis 5 zeigen denselben Buchstaben, wie er in den drei Größen 8-Point, 10-Point und 12-Point (ein "Point" = 1/72 Zoll) auf einem handelsüblichen Laserdrucker mit einer Punktdichte von 300 · 300 Punkten pro Quadratzoll ausgegeben werden kann; im Original entsprechen diese drei Matrizes etwa den Druckgrößen א, א und א.

Die Erstellung der notwendigen Zeichensätze (sog. "Fonts") ist also der erste zu bewältigende Schritt bei der Entwicklung. Allerdings können die verschiedenen Größen nur unter großem Qualitätsverlust automatisch auseinander abgeleitet werden. Hier zeichnet sich für die Zukunft eine Vereinfachung ab, insofern für Laserdrucker anstelle von Bitmatrixfonts mehr und mehr auf sogenannte Outlinefonts übergegangen wird, bei denen nicht mehr einzelne Punkte, sondern die Umrisse der Zeichen als Linien mit Anfangs- und Endpunkten definiert werden, die der Drucker durch Umrechnung dann selbständig auf die gewünschte Größe bringt (vgl. Abbildung 6 mit einer Outline-Darstellung des hebräischen Aleph). Sowohl für die Generierung über Matrizes als auch für die von Outlinefonts stehen heute kommerzielle Programme zur Verfügung.





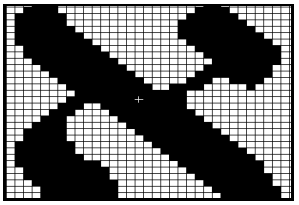
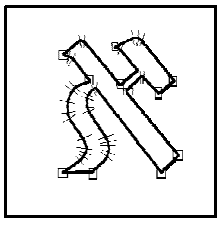


Abb. 3 bis 5: Bitmatrizes für Laserdrucker (8-Pt. / 10-Pt. / 12-Pt.-Größe)

Eine zweite Entwicklungseinheit betrifft die Verwaltung der erstellten Zeichen durch den Rechner. Hier gibt es grundsätzlich zwei konkurrierende Verfahren, den sog. Text- oder Alphamodus und den sog. Grafikmodus. Beide unterscheiden sich im wesentlichen dadurch, daß im ersteren Fall zunächst der vollständige Zeichensatz in den für den Bildschirm vorgesehenen Speicher (sog. "Bildschirm- oder Grafikkarte") bzw. in den Speicher des Druckers geladen wird (sog. "Downloadverfahren") und die Geräte daraufhin über bestimmte Codes angewiesen werden, die Matrix des anzusteuernden Zeichens aus diesem Zeichensatz herauszulesen und auszugeben; im Grafikmodus wird hingegen die erforderliche Punktmatrix für jedes einzelne darzustellende Zeichen vom Rechner an die Ausgabegeräte weitergeleitet. Das letztere Verfahren ist wesentlich flexibler, da die Anzahl der gleichzeitig verwendbaren Zeichensätze hier praktisch unbegrenzt ist, während sie im Downloadverfahren von der Kapazität des Druckers bzw. der Bildschirmkarte abhängt. Andererseits erlaubt der Textmodus im Normalfall eine wesentlich schnellere Verarbeitung, da der für die Definition eines Zeichens benötigte, bei größeren Matrizes recht umfangreiche Code nur einmal ausgegeben zu werden braucht und dann über einen wesentlich kürzeren Code abrufbar ist. Die Verwaltung der Zeichensätze und der Steuercodes ist die zentrale Funktion der sog. Bildschirm- und Druckertreiber.

Abb. 6: Hebräisches Aleph in Outline-Darstellung



Eine am gestalterischen Optimum orientierte Verarbeitung von Schriften wird nicht nur bei Lateinschriften, sondern auch bei den meisten anderen Schriften von den bei Schreibmaschinen üblichen festen Schrittabständen abgehen und stattdessen eine "proportionale" Wiedergabe vorziehen, bei der jedem einzelnen Zeichen die seiner Form entsprechende Breite zukommt; man vgl. etwa die Lateinbuchstaben *i* und *m*, die sich in ihrer natürlichen Breite etwa um einen Faktor drei unterscheiden. Einem Programm, das die optimale Füllung einer Zeile oder auch die nötigen Wortzwischenräume für die Erstellung von Blocksatz errechnen soll, muß die spezifische Weite eines jeden Zeichens bekannt sein; diese Aufgabe übernehmen die sog. "Weitentabellen", die eine eigene zu erstellende Treibereinheit bilden.

Die wohl aufwendigste Treibereinheit betrifft die Ansteuerung der gewünschten Zeichen über die Tastatur. Prinzipiell ist davon auszugehen, daß jede Taste auf der Tastatur eines Computers beim Niederdrücken einen bestimmten elektronischen Code erzeugt, der vom Computer empfangen und interpretiert wird. Daß beim Niederdrücken der Taste "M" tatsächlich ein "M" erzeugt wird, ist nicht selbstverständlich und beruht ausschließlich auf der Interpretation des Tastencodes durch den Rechner. Durch ein eigenes Programm kann der Rechner nun angewiesen werden, aufgrund desselben Tastencodes einmal ein "M" und einmal ein arabisches م zu erzeugen. Je nach der Menge und Art der zu verarbeitenden Einzelzeichen einer Schrift und ihrer internen Codierung durch das Textverarbeitungsprogramm ist die Gestaltung eines solchen Tastaturtreibers mehr oder weniger komplex, wobei im orientalischen Bereich die Sonderproblematik von linksläufigen Schriften wie der arabischen, vertikal angeordneten Schriften wie der klassisch-mongolischen, von Silbenschriften wie der koreanischen oder von Wortschriften wie der chinesischen zu berücksichtigen ist. Pauschale Lösungen gibt es hier nicht; vielmehr benötigt jede Schrift ihre eigene Anpassung, die vom Ideal eines möglichst einfachen Eingabe- und Umschaltmodus geprägt sein sollte.

Anzustreben ist nach alledem eine Textverarbeitung, die den gesamten Bereich orientalischer Schriften abdeckt. Die Realisierung eines solchen Vorhabens, an der die Universität Bamberg durch die neugeschaffene Arbeitsstelle für Orientalistische Computerlinguistik beteiligt ist, wird noch einige Zeit in Anspruch nehmen.

Die Verwendung des Computers in der Orientalistik braucht sich aber nicht auf die Erzeugung von Manuskripten oder Druckvorlagen, die Originalschriften enthalten, zu beschränken. Mehr und mehr wird der Computer hier, wie bereits in anderen geisteswissenschaftlichen Disziplinen, als Hilfsmittel bei der Analyse und Auswertung Verwendung finden. Dabei ist zunächst wieder an den Bereich sprachlicher Daten zu denken, wobei etwa im lexikographischen Sektor die elektronische Erstellung von Wortindizes und Textkonkordanzen als Grundlage für die Erarbeitung von Wörterbüchern in Betracht kommt. Voraussetzung hierfür ist zunächst die Erfassung von Texten, wobei prinzipiell zwei Verfahren anwendbar sind: Als die "klassische" Methode die Eingabe von Hand und als zukunftsträchtige Alternative das automatische Einlesen per Scanner.

Daß die Orientalistik auch in dieser Hinsicht gegenüber anderen philologischen Fächern im Rückstand ist, beruht wiederum auf der Problematik der Originalschriften. Das betrifft weniger die manuelle Eingabe, da hier auf Codierungen zurückgegriffen werden kann, die im Sinne einer provisorischen Transkription sogar mit den reinen "ASCII-Zeichen" auskommen können. Mit einer solchen Codierung sind in den USA bereits vor Jahren z.B. die hebräische Bibel oder der altindische Ṛgveda eingespeichert worden. Soweit die Codierung dabei eindeutig war, lassen sich diese Texte heute völlig problemlos in originalschriftliche oder transkriptive Codierungen konvertieren; man vgl. etwa die folgende Passage aus dem Alten Testament (Hos. 9,1-5):
Ein Problem stellen orientalische Originalschriften hingegen in jeglicher Hinsicht bei der Texterfassung mittels Scanner dar. Handelsübliche Texterkennungssoftware ist grundsätzlich auf die Gegebenheiten lateinschriftlicher Vorlagen zugeschnitten. Dabei wird das eingelesene Abbild einer Vorlagenseite zunächst auf den Freiraum zwischen den Zeilen hin durchsucht; innerhalb der so gewonnenen Zeilen werden die einzelnen Buchstaben durch den Abstand voneinander isoliert und durch den Vergleich mit vorgegebenen Mustern identifiziert. Sofern die Schrifttype einheitlich ist und die Druckqualität nur geringe Abweichungen von den Erkennungsmustern zeigt, ist hier bereits eine hohe Genauigkeitsstufe erreichbar. Der Versuch einer Anpassung an andere Schriften ist bei diesem Verfahren aber nur dann erfolgversprechend, wenn diese nach demselben Prinzip wie die Lateinschrift, nämlich als Buchstabenschriften organisiert sind. Das gilt im Bereich der Orientalistik z.B. für die georgische Schrift, die sich aufgrund ihrer Ausprägung recht gut für ein automatisches Einlesen eignet, wobei allerdings die Ähnlichkeit gewisser Buchstaben wie z.B. ბ = *b* und გ = *g*, ო = *o* und თ = *t*, ღ = *ġ* (spirantisches *g*) und დ = *d* oder ვ = *v*, კ = *ḳ* und ჳ = *w* trotz sorgfältiger Eingabe von Erkennungsmustern ("Teaching") immer wieder zu Verwechslungen führt; man vgl. etwa das folgende Beispiel, bei dem für das eingelesene Druckstück (Hos. 9,1-5) eine transkriptive Codierung gewählt wurde (durch * sind nicht erkannte Buchstaben gekennzeichnet, verwechselte Buchstaben sind fett und unterstrichen dargestellt):

Taf. 1: Hos. 9,1-5 in der Codierung der University of Pennsylvania (CCAT)

Ausgangstext:

9:1)AL-T.I&:MA63X YI&:RF)"70L05)EL-G.IYL03 K.F75/(AM.I8-0YM?K.I71Y ZFNI73YTF M"/(A74L):ELOHE92Y/KF?)FHA74-B:T.F)ET:NF80N (A73L K.FL- G.FR:NO71WT D.FGF75N00?
2 G.O71REN WF/YE73QEB LO74) YIR:(/"92M W:/TIYRO73W$ Y:KA71XE$ B./F75H.00?
3 LO71) Y"$:B73W. B.:/)E74REC Y:HWF92H?W:/$F70B)EP:-RA63YIM03 MIC:RA80YIM W./B:/)A$.73W.R +FM"71) YO)-K"75LW.00?
4 LO)-YIS.:K63W. LA/YHWF71H05 YAYIN02 W:/LO74) YE75-(ER:BW.L/OW01ZIB:X"Y/HE81M?K.:/LE70XEM)OWNIYM03L/F-HE80M K.FL-)OK:LF73Y/W YI+AM.F92)W.?K.I75Y-LAX:M/F74M L:/NAP:$/F80M LO71) YFBO73W) B."71YT Y:HWI75H00?
5 MA75H-T.A(:A&73W. L:/YO74WM MOW("92D W./L:/YO73WM XAG-Y:HWF75H00?

Transkription:

9:1 *ʾal-tiśmaḥ yiśrāʾēl ʾel-gīl kāʾammīm kī zānîtā mēʿal ʾélohêkā ʾāhabtā ʾetnān ʿal kål-gārnōt dāgan;*
2 *gorēn wāyeqeb loʾ yirʿēm wə-tîrôš yəkaheš bāhh;*
3 *loʾ ʿyēšbū bəʾereṣ ʾadonay wəšāb ʾepʳayim miṣrayim ūbə-ʾaššūr ṭāmeʾ yoʾkelū;*
4 *loʾ-yisskū laʾadonay yayin wəloʾ yeʿerbū-lō zībhèhem kə-lehem ʾōnîm lāhem kål-ʾoklāyw yiṭammāʾū kī-lahmām lənapšām loʾ yābōʾ bèt adonay.*
5 *māh-taʿāśū ləyōm môʿēd ūləyōm ḥag-adonay;*

Originalschrift:

ט:1 אַל־תִּשְׂמַח יִשְׂרָאֵל אֶל־גִּיל
כָּעַמִּים כִּי זָנִיתָ מֵעַל אֱלֹהֶיךָ אָהַבְתָּ
אֶתְנָן עַל כָּל־גָּרְנוֹת דָּגָן׃
2 גֹּרֶן וְיֶקֶב לֹא יִרְעֵם וְתִירוֹשׁ
יְכַחֲשׁ בָּהּ׃
3 לֹא יֵשְׁבוּ בְּאֶרֶץ יְהוָה וְשָׁב אֶפְרַיִם
מִיצְרַיִם וּבְאַשּׁוּר טָמֵא יֹאכֵלוּ׃
4 לֹא־יִסְכּוּ לַיהוָה יַיִן וְלֹא
יֶעַרְבוּ־לוֹ זִבְחֵיהֶם כְּלֶחֶם אוֹנִים
לָהֶם כָּל־אֹכְלָיו יִטַּמָּאוּ כִּי־לַחְמָם
לְנַפְשָׁם לֹא יָבוֹא בֵּית יְהוָה׃
5 מַה־תַּעֲשׂוּ לְיוֹם מוֹעֵד וּלְיוֹם
חַג־יְהוָה׃

Taf. 2: Hos. 9,1-5 aus der georgischen Mcxeta-Bibel

Original:

1. ნუ იხარებ, ისრაჱლ, ნუ იშუებ, ვითარცა ერნი შენნი, რამეთუ განჰმეძენ ღმრთისაგან შენისა. შეიყუარენ საცემელნი ყოველთა ზედა კალოსა ოფჭლისასა.
2. კალომან და საწნეხელმან არა იცნნა იგინი და ღვნომან უტყუა მათ.
3. არა დაემკუდრნეს ქუეყანასა შინა უფლისასა. დაემკუდრა ეფრემ ეგჳპტესა და ასსურასტანელთა შორის, არაწმიდათა ჭამდენ,
4. არ აღუფჭურიეს უფალსა ღუნოჲ და არა დაუტკბნეს მათ მსხუერპლნი მათნი, ვითარცა პურთა გლოისა მათისათა. ყოველნი მჭამელნი მათნი შეიგინნენ მით, რამეთუ პურნი მათნი სულთა მათთანი არა შევიდენ სახიდ უფლისა.
5. რაჲ ჰყო დღესა შინა კრებისასა და დღესა შინა დღესასწაულისა უფლისასა?

Taf. 3: Dasselbe nach Einlesen mit Scanner und Konvertierung

Leseresultat:

1. *n* ixareb, israel, nu išueb, vitarca erni šenni, rametu ganhmeʒven *mrtisa**b**an šenisa. šeiquaren sacemelni ǫovelta zeda ḳal*sa ifklisasa.*
2. *ḳaloman da saqnexelman ara icnna igini da ġwnoman utąua mat.*
3. *ara daemḳwdrnes kueąanasa šina uplisasa. daemḳwdra eprem egwṗtesa da assu*asṭanelta šoris, araqmidata čamden,*
4. *ar agupkuries upalsa ġwnoy da ara dau*ḳbnes mat msxverṗlni matni, vitarca ṗurta gloisa matisata. ǫovelni mčamelni matni šeiginnen mit, rame**o**u ṗurni matni sulta mattani ara ševiden saxid uplisa.*
5. *ray hąo d**d**esa šina ḳrebi*asa da dġesa šina dġesasqaulisa uplisasa?*

Die meisten anderen orientalischen Schriften wie z.B. die arabische verlangen demgegenüber ein abweichendes Verfahren, bei dem eine andere Abgrenzungsbasis als der Abstand zwischen einzelnen Zeichen zu wählen ist; die Entwicklung solcher Verfahren ist als ein vordringliches Desiderat einzustufen.

Ein weiteres Desiderat im Hinblick auf den Einsatz des Computers bei der sprachwissenschaftlichen Analyse orientalischer Textmaterialien sind Verfahren für eine morphologische Sortierung. Obwohl schon heute durchaus leistungsfähige Programme vorliegen, die Wortformenindizes und Textstellenkonkordanzen aus elektronisch gespeicherten Texten ableiten, reichen die erzielten Resultate für wissenschaftliche Fragestellungen häufig nicht aus, da das Ordnungsprinzip im Normalfall eine Auflistung nach dem Alphabet der Wortformen ist; um ein benutzbares Lexikon zu schaffen, wird hingegen eine Lemmatisierungsfunktion benötigt, die von suffixalen oder sogar präfixalen Elementen abstrahiert und Flexionsformen in ihren paradigmatischen Zusammenhang einordnet. Man vergleiche dazu z.B. die folgenden Auszüge aus einem automatisch erstellten Index und einer ebenso erzeugten Konkordanz zu dem etwa 300 Seiten umfassenden Sammelband von Volksliedtexten in der Sprache der kaukasischen Svanen (Svanuri Ṗoezia, Tbilisi 1939):

Taf. 4: Wortformenindex zu Svanuri Ṗoezia (Auszug)

Wortformenindex zu Svanuri Ṗoezia:

(In Klammern: Die Häufigkeit der betreffenden Wortform)

abaz	(1)	110h: 350, 4
abram	(2)	64b: 236, 82; 66: 242, 25
abrams	(1)	66: 240, 3
abrešumiš	(1)	1b: 4, 13
abrešvimiš	(1)	50a: 174, 24
abžinalix	(1)	5: 18, 46
abǯare	(1)	1a: 2, 7
abǯari	(1)	29: 100, 7
abǯaris	(2)	32: 110, 14; 41b: 136, 14
abǯriš	(1)	41b: 138, 52
acar	(1)	66: 242, 12
acars	(2)	66: 240, 7; 242, 15
aceri	(1)	66: 242, 20
acurax	(1)	94a: 290, 36
acvir	(2)	31: 108, 65; 43b: 154, 64
acvird	(1)	20: 66, 37
aċhanyeli	(1)	30: 102, 2
aċhi	(1)	67: 244, 27
ačad	(18)	8: 30, 102; 18: 62, 13; 27a: 92, 73; 41a: 132, 9; 41b: 136, 9.11; 43b: 152, 39; 77a: 256, 1.3.4.5.6. 7.8; 258, 11.12; 77b: 258, 4; 91b: 270, 18
ačadd	(1)	57a: 190, 37
ačadx	(2)	9: 38, 56; 94a: 294, 118
ače	(2)	28: 98, 22.33
ačed	(1)	39b: 124, 5
ačkad	(2)	22: 70, 15.17
ačunġo	(1)	26: 88, 72

Textstellenkonkordanz zu Svanuri Ṗoezia:

			xoqaci (1)
72/Mlx:	(248),6	*dosgu*	*xoqaci macxvari!*
			xoqde (1)
43a/Lšx:	(150),30	*merma ṭatxas gvalvars*	*xoqde".*
			xoqdex (2)
25b/Mlx:	(80),25	*xexvas mineš ečav*	*xoqdex;*
	(80),27	*begärs mine šənšv ečav*	*xoqdex;*
			xoqida (4)
8/Mlx:	(28),79	*päsild uṭkläbvd ka*	*xoqida.*
9/Mṣṭ:	(38),50	*ǯveg i meǯveg mäg*	*xoqīda.*
41b/Ḳal:	(134),7	*baṗəld ägite*	*xoqida,*
46/Mlx:	(166),107	*mišgov lerekv ka*	*xoqīda;*
			xoqidax (5)
8/Mlx:	(26),54	*atxa kämte ċur*	*xoqidax,*
13/Mlx:	(46),15	*sga*	*xoqīdax, qän sga xobax,*
23/Lǯr:	(70),5	*ži*	*xoqidax zagärteži.*
51/Lṭl:	(164a),68	*sgav*	*xoqidax Qarāštēsga,*
72/Mlx:	(248),2	*ǯvegi*	*xoqidax didi mindvriše.*

Taf. 5: Textstellenkonkordanz zu Svanuri Ṗoezia (Auszug)

Obwohl auch solche rein alphabetischen Auflistungen bereits einen eigenen Wert haben, da sie als Grundstock für eine vollständige Erfassung der auftretenden Wortformen und darauf aufbauende morphologische und phonologische Analysen dienen können, wäre eine weitergehende Nutzung erst dann möglich, wenn die paradigmatische Zusammengehörigkeit von Wortformen wie *ačad* und *otčed* (beides finite Formen einer Verbalwurzel *-čed-* "gehen") oder von *xoqde* und *miqida* (beides finite Formen einer Verbalwurzel *-qed-* "gehen") erkannt würde und als übergelagertes Ordnungsprinzip eingesetzt wäre. Eine solche Funktion muß natürlich die internen grammatischen Regeln der jeweiligen Objektsprache reflektieren und kann nicht ohne weiteres auf pauschale Algorithmen zurückgreifen. Auch hierzu bedarf es weiterführender Entwicklungen, an denen sich die Arbeitsstelle Orientalistische Computerlinguistik der Universität Bamberg in Kooperation mit anderen in- und ausländischen Hochschulen beteiligen wird.